Stimulus-dependent maximum entropy models of neural population codes

Einat Granot-Atedgi[=,a], Gašper Tkačik[*=,b], Ronen Segev[≡,c], and Elad Schneidman[≡,a]
[a] *Department of Neurobiology, Weizmann Institute of Science, 76100 Rehovot, Israel*
[b] *Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria*
[c] *Facutly of Natural Sciences, Department of Life Sciences and Zlotowski Center for Neuroscience, Ben Gurion University of the Negev, 84105 Be'er Sheva, Israel*

(Dated: March 5, 2013)

Neural populations encode information about their stimulus in a collective fashion, by joint activity patterns of spiking and silence. A full account of this mapping from stimulus to neural activity is given by the conditional probability distribution over neural codewords given the sensory input. To be able to infer a model for this distribution from large-scale neural recordings, we introduce a stimulus-dependent maximum entropy (SDME) model—a minimal extension of the canonical linear-nonlinear model of a single neuron, to a pairwise-coupled neural population. The model is able to capture the single-cell response properties as well as the correlations in neural spiking due to shared stimulus and due to effective neuron-to-neuron connections. Here we show that in a population of 100 retinal ganglion cells in the salamander retina responding to temporal white-noise stimuli, dependencies between cells play an important encoding role. As a result, the SDME model gives a more accurate account of single cell responses and in particular outperforms uncoupled models in reproducing the distributions of codewords emitted in response to a stimulus. We show how the SDME model, in conjunction with static maximum entropy models of population vocabulary, can be used to estimate information-theoretic quantities like surprise and information transmission in a neural population.

INTRODUCTION

Neurons represent and transmit information using temporal sequences of short stereotyped bursts of electrical activity, or spikes [1]. Much of what we know about this encoding has been learned by studying the mapping between stimuli and responses at the level of single neurons, and building detailed models of what stimulus features drive a single neuron to spike [2–4]. In most of the nervous system, however, information is represented by joint activity patterns of spiking and silence over populations of cells. In a sensory context, these patterns can be thought of as codewords that convey information about external stimuli to the central nervous system. One of the challenges of neuroscience is to understand the neural *codebook*—a map from the stimuli to the neural codewords—a task made difficult by the fact that neurons respond to the stimulus neither deterministically nor independently.

The structure of correlations among the neurons determines the organization of the code, that is, how different stimuli are represented by the population activity [5–8]. These correlations also determine what the brain, having no access to the stimulus apart from the spikes coming from the sensory periphery, can learn about the outside world [9–11]. The source of these correlations, which arise either from the correlated external stimuli to the neurons, from "shared" local input from other neurons, or from "private" independent noise, have been heavily debated [12–15]. In many neural systems, the correlation between pairs of (even nearby or functionally similar) neurons was found to be weak [16, 17, 26]. Similarly, the redundancy between pairs in terms of the information they convey about their stimuli was also typically weak [18–20]. The low correlations and redundancies between pairs of neurons therefore led to the suggestion that neurons in larger populations might encode information independently [21], which was echoed by theoretical ideas of maximally efficient neural codes [22–24].

Recent studies of the neural code in large populations have, however, revealed that while the typical pairwise correlations may be weak, larger populations of neurons can nevertheless be strongly correlated as a whole [25–33]. Maximum entropy models of neural populations have shown that such strong network correlations can be the result of collective effects of pairwise dependencies between cells, and, in some cases, of sparse high-order dependencies [26, 34, 35]. Most of these studies have characterized the strength of network effects and spiking synchrony at the level of the total *vocabulary* of the population, i.e. the distribution of codewords averaged over all the stimuli. It is not immediately clear how these findings affect stimulus encoding, where one needs to distinguish the impact of correlated stimuli that the cells receive ("stimulus correlations"), from the impact of covariance of the cells conditional on the stimulus ("noise correlations"). For small populations of neurons, it has been shown that taking into account correlations for decoding or reconstructing the stimulus can be beneficial compared to the case where correlations are neglected (e.g. [35–39]). Similarly, generalized linear models high-

[*] gtkacik@ist.ac.at
=,≡ equal contributions

lighted the importance of dependencies between cells in accounting for correlations between pairs and triplets of retinal ganglion cell responses [40].

Here we present a new encoding model that allows us to study in fine detail the codebook of large neural populations. Importantly, this model gives a joint probability distribution over the activity patterns of the whole population for a given stimulus, while capturing both the stimulus and noise correlations. This new model belongs to a class of maximum entropy models with strong links to statistical physics [27, 41–52] and is directly related to maximum entropy models of neural vocabulary [26–32], allowing us estimate the entropy and its derivative quantities for the neural code. In sum, the maximum entropy framework enables us to progress towards our goal of focusing attention on the level of joint patterns of activity, rather than capturing low-level statistics (e.g., the individual firing rates) of the neural code alone.

We start by showing that linear-nonlinear (LN) models of retinal ganglion cells responding to spatially unstructured stimuli capture a significant part of the single neuron response, but still miss much of the detail; in particular, we show that they fail to capture the correlation structure of firing among the cells. We next present our new *stimulus-dependent maximum entropy* (SDME) model, which is a hybrid between linear-nonlinear models for single cells and the pairwise maximum entropy models. Applied to groups of ~ 100 neurons recorded simultaneously, we find that SDME models outperform the LN models for the stimulus-response mapping of single cells and, crucially, give a significantly better account of the distribution of codewords in the neural population.

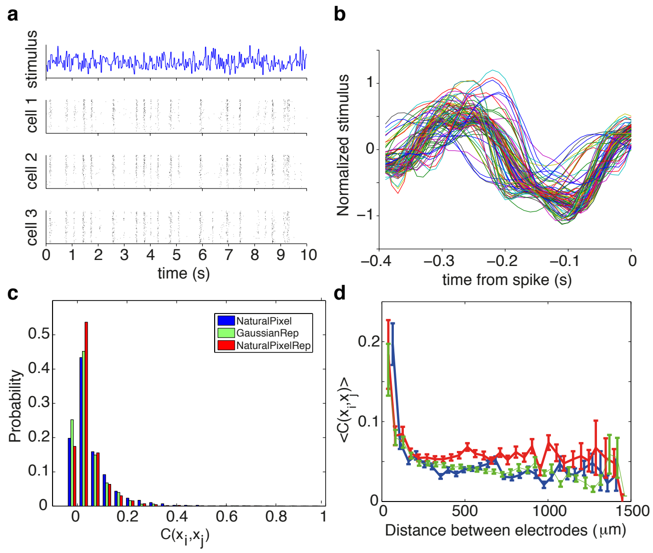

FIG. 1: **Response of a large population of ganglion cells to a 10 s long repeated visual stimulus**. **(a)** White noise uncorrelated Gaussian stimulus presented at 30 Hz and the spiking patterns of 3 cells to repeated presentations of the stimulus. **(b)** Spike-trigerred averages of 110 simultaneously recorded cells; a subset of 100 cells was chosen for further analysis. **(c)** The histogram of pairwise correlations between cells for repeated Gaussian white noise stimulus (green), repeated natural pixel movie (red), and non-repeated natural pixel movie (blue) [35]. **(d)** Average pairwise correlation coefficient between cells as a function of the distance (mean and std are across pairs of cells at a given distance).

RESULTS

We recorded the simultaneous spiking activity of ~ 110 ganglion cells from the salamander retina [53], presented with repeats of a 10 s long full-field flicker ("Gaussian FFF") movie, where the light intensity on the screen was sampled independently from a Gaussian distribution with a frequency of 30 Hz (Fig. 1a). This "frozen noise" stimulus was repeated 726 times, for a total of ~ 2 h of stimulation. Most of the recorded cells exhibited temporal OFF-like behaviors (Fig. 1b). We chose for further analysis $N = 100$ cells that were reliably sorted, demonstrated a robust and stable response over repeats, and generated at least 2500 spikes during the course of the experiment.

We discretized neural responses into $\Delta t = 10$ ms bins, and represented the activity of the neurons in response to the stimulus as binary words in each of the time bins. If neuron $i = 1, \dots, N$ was active in time bin t, we denoted a spike (or more spikes) as $x_i(t) = 1$, and $x_i(t) = 0$ if it was silent. In this representation, the whole experiment yielded a total of about $T \sim 7.3 \cdot 10^5$ binary word samples. Using repeated presentations of the same movie, we estimated the average response of each of the cells across repeats, $r_i(t) = \langle x_i(t) \rangle_{rep}$, or the peri-stimulus time histogram (PSTH). Following Refs. [4, 54], we fitted a linear-nonlinear model for each of the cells in the experiment, such that the predicted rate $r_i^{LN}(t) = \mathcal{N}_i(\mathbf{k}_i \cdot \mathbf{s}(t))$, where $\mathbf{k}_i$ is a linear filter matched for the i-th cell, $\mathcal{N}_i$ is its point-wise nonlinear function, and $\mathbf{s}(t)$ is the stimulus fragment from time $t - \tau$ until t (here we used $\tau = 400$ ms, making $\mathbf{s}(t)$ a vector of light intensities with 40 components). Linear filters were reconstructed using reverse correlation (spike-triggered average), and nonlinearities were obtained by histograming $P(\mathbf{k}_i \cdot \mathbf{s}(t)|\text{spike})$ into $K = 20$ adaptively-sized bins and obtaining $r_i^{LN}(t) = \mathcal{N}_i(\mathbf{k}_i \cdot \mathbf{s}) = P(\text{spike}|\mathbf{k}_i \cdot \mathbf{s}(t))$ by inverting $P(\mathbf{k}_i \cdot \mathbf{s}(t)|\text{spike})$ using Bayes' rule. These LN models captured most of structure of the PSTH, yet as the example cell in Fig. 2a shows, they often misestimated the exact firing rates of the neuron, or sometimes even missed parts of the neural response altogether. In the Gaussian FFF condition, the normalized (Pearson) correlation between the measured and predicted PSTH, $\text{Corr}(r_i(t), r_i^{LN}(t))$, was 0.69 ± 0.06 (mean ± std across 100 cells).

The performance gap of the canonical LN models in predicting single neuron responses suggests that either the single-neuron models need to be improved to account for the observed behavior, or that interactions between neurons play an important encoding role and need to be included. While firing rate prediction performance can be improved for single neurons by models with higher-dimensional stimulus sensitivity (e.g. [54, 55]) or dynamical aspects of spiking behavior (e.g. [56, 57]), previous work, as well as the results below, demonstrated that even conditionally-independent models which by construction perfectly reproduce the firing rate behavior of single cells, often fail to capture the measured correlation structure of firing between pairs of cells, as well as higher-order statistical structure [26].

We find two salient features of the correlations between pairs of neurons: (i) the pairwise correlations between cells in response to the Gaussian FFF movie are typically weak but are not zero (Fig. 1c, consistently with previous reports [26, 27, 32]); (ii) the correlation in neural activities shows a fast decay with distance despite the infinite correlation length of the stimulus, but the decay does not reach zero correlation even at relatively large distances (Fig. 1d). This salient structure, along with any other potential statistical correlation at the pairwise order, is characterized by the covariance matrix of activities, $C_{ij} = \langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle$, where the averages are taken across time and repeats.

We would like to find a model of the neural code that will be able to reproduce these pairwise statistics. Motivated by the shortcomings of the canonical LN model, we therefore asked whether a model that combined the LN (receptive-field based) aspect of single cells with the interactions between cells, could give a better account of the neural stimulus-response mapping. Importantly, the new model should capture not only the firing rate of single cells and the full pairwise correlation structure between them, but should also accurately predict the full distribution of the joint activity patterns across the whole population. Because the joint distributions of activity are high-dimensional (e.g., the distribution over codewords across the duration of the experiment, $P(\{x_i\})$, has 2^N components), this is a very demanding benchmark for any model.

Here we propose the simplest extension to the conditionally-independent set of LN models for each cell in the recorded population, by including pairwise couplings between cells, so that the spiking of cell i can increase or decrease the probability of spiking for cell j [58, 59]. In contrast to previous proposals, this coupling will be introduced so that the resulting model is a maximum-entropy model for $P(\{x_i\}|\mathbf{s})$, the conditional distribution over population activity patterns given the stimulus. We recall that the maximum entropy models give the most parsimonious probabilistic description of the joint activity patterns, which perfectly reproduces a chosen set of measured statistics over these patterns, without making any additional assumptions [60].

We start by introducing the least structured (maximum entropy) distribution $P(x_1, x_2, \ldots, x_N|t)$ that reproduces exactly the observed average firing rate for each time bin t and for each neuron i, $r_i(t) = \langle x_i(t)\rangle_{data} = \langle x_i(t)\rangle_P$, as well as the overall correlation matrix C_{ij} between all pairs of cells (c.f. [61]). Thus, we seek $P(\{x_i\}|t)$ that maximizes $\mathcal{L}$:

$$\begin{aligned}\mathcal{L}\left[P(\{x_i\}|t)\right] &= -\sum_{\{x_i\},t} P(\{x_i\}|t)\log_2 P(\{x_i\}|t)\\ &+ \sum_{i,t}\alpha_i(t)[\langle x_i(t)\rangle_P - \langle x_i(t)\rangle_{data}]\\ &+ \frac{1}{2}\sum_{ij}\beta_{ij}[\langle x_i x_j\rangle_{P,t} - \langle x_i x_j\rangle_{data}]\\ &+ \sum_{\{x_i\},t}\mu(t)[P(\{x_i\}|t) - 1],\end{aligned} \tag{1}$$

where the subscript to brackets $\langle\cdot\rangle$ denotes whether the averaging is done over the maximum entropy distribution (P), or over the recorded data; Lagrange multipliers μ ensure that the distributions are normalized. This is an optimization problem for parameters $\alpha_i(t)$ and β_{ij}, which has a unique solution since the entropy is convex. The functional form of the solution to this optimization problem is well-known; in our case it can be written as

$$\begin{aligned}&P^{TDME}(\{x_i\}|t) = \\ &\frac{1}{Z(t)}\exp\left(\sum_{i=1}^{N}\alpha_i(t)x_i + \frac{1}{2}\sum_{i,j=1}^{N}\beta_{ij}x_i x_j\right),\end{aligned} \tag{2}$$

where the individual time-dependent parameters for each of the cells, $\alpha_i(t)$, and the stimulus-independent pairwise interaction terms β_{ij}, are set to match the measured firing rates $r_i(t)$ and the pairwise correlations C_{ij}; $Z(t)$ is a normalization factor or partition function for each time bin t, given by $Z(t) = \sum_{\{x_i\}}\exp\left(\sum_i \alpha_i(t)x_i + \frac{1}{2}\sum_{ij}\beta_{ij}x_i x_j\right)$.

The time-dependent maximum entropy (TDME) model in Eq. (2) is equivalent to an Ising model from physics, where the single-cell parameters are time-dependent local fields acting on each of the neurons (spins), and static (stimulus-independent) infinite-range interaction terms couple each pair of spins. In the limit where interactions go to zero, $\beta_{ij} = 0$, the model in Eq. (2) becomes the full conditionally-independent model, itself a maximum entropy model that reproduces exactly the firing rate of every neuron, $r_i(t)$; in this case the probability distribution factorizes, and the solution for $\alpha_i(t)$ and $Z(t)$ becomes trivially computable from the firing rates, $r_i(t)$. Time-dependent maximum entropy models are powerful, since they make no assumptions about how the stimulus drives the response; they can serve as useful benchmarks for other models (especially the conditionally independent model with $\beta_{ij} = 0$). On the other hand, these models require repeated stimulus presentations to fit, involve a number of parameters that grows linearly with the duration of the stimulus, do not generalize to new stimuli, and do not provide an explicit map from the stimuli to the responses.

To make a direct link to the stimulus and allow comparison with a set of uncoupled LN models, we take the general time-dependent maximum entropy model of Eq. (2) and specialize it to a particular form of stimulus dependence. Rather than having an arbitrary time-dependent parameter for every neuron for each time bin, $\alpha_i(t)$, we assume that this dependence takes place through the stimulus projection alone, i.e. $\alpha_i(t) = \alpha_i(\mathbf{k}_i \cdot \mathbf{s}(t))$, much like in an LN model, where the neural firing depends on the value of the stimulus projection onto the linear filter $\mathbf{k}_i$. This choice is made purely for the sake of convenience: the model could be generalized to, e.g., neurons that depend on two linear projections of the stimulus, by making α_i depend jointly on $(\mathbf{k}_1 \cdot \mathbf{s}(t), \mathbf{k}_2 \cdot \mathbf{s}(t))$, although such models would be progressively more difficult to infer from data.

Concretely, we estimated the linear filter $\mathbf{k}_i$ for each cell i using reverse correlation, and convolved the filter with the stimulus sequence, $s(t)$, to get the "generator signal" $g_i(t) = \mathbf{k}_i \cdot \mathbf{s}(t)$. We then looked for the maximum entropy probability distribution $P(\{x_i\}|\mathbf{s}(t))$, by requiring that the average firing rate of every cell given the generator signal is the same in the data and under the model, i.e. $\langle x_i(g_i)\rangle_{data} = \langle x_i(g_i)\rangle_P$ (see Methods); as before, we also required that the model reproduced the overall correlation between every pair of cells, C_{ij}. This gives then a stimulus-dependent maximum entropy (SDME) model, which takes the following form:

$$P^{SDME}(\{x_i\}|\mathbf{s}(t)) = \qquad (3)$$
$$\frac{1}{Z(\mathbf{s}(t))} \exp\left(\sum_{i=1}^{N} \alpha_i(g_i(t))x_i + \frac{1}{2}\sum_{i,j=1}^{N} \beta_{ij}x_ix_j\right).$$

The parameters of this model are: $N \times (N-1)/2$ couplings β_{ij}, $K \times N$ parameters α_i, and a linear filter $\mathbf{k}_i$ for each cell. We used a Monte Carlo based gradient descent learning procedure to find the model parameters α, β numerically (see Methods).

By construction, the SDME model exactly reproduces the covariance in activities, C_{ij}, between all pairs of cells, and also the LN model properties of every cell: an arbitrary nonlinear function $\mathcal{N}$ can be encoded by properly choosing how parameters α_i depend on the linear projections of the stimulus, g_i. We can construct a maximum entropy model with $\beta_{ij} = 0$ (no constraints on the pairwise correlations C_{ij}). The result is a set of uncoupled (conditionally independent) LN models:

$$P^{LN}(\{x_i\}|\mathbf{s}(t)) \equiv \prod_{i=1}^{N} \frac{1}{\tilde{Z}_i(\mathbf{s}(t))} \exp\left(\tilde{\alpha}_i(g_i(t))x_i\right) \qquad (4)$$
$$= \prod_{i=1}^{N} \mathcal{N}_i(\mathbf{k}_i \cdot \mathbf{s}(t)).$$

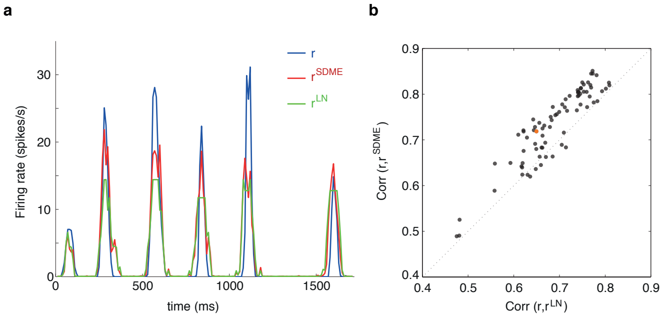

FIG. 2: **SDME model predicts the firing rate of single cells better than LN models.** **(a)** Example of the PSTH segment for one cell (blue), the best prediction of an LN model (green) and of a SDME model (red). **(b)** Correlation between the true PSTH and SDME model prediction (vertical axis) vs. the correlation between the true PSTH and the LN model prediction (horizontal axis); each plot symbol is a separate cell, dotted line shows equality. The neuron chosen in panel (a) is shown in orange.

In sum, the time-dependent maximum entropy (TDME) model of Eq. (2) is an extension of conditionally independent model that additionally reproduces the measured pairwise correlations between cells. In a directly analogous way, the stimulus-dependent maximum entropy (SDME) model of Eq. (3) is an extension to the set of uncoupled LN models, Eq. (4), that additionally reproduces the measured pairwise correlations between cells. Because P^{SDME} (Eq. 3) agrees with P^{LN} (Eq. 4) exactly in all constrained single-neuron statistics, any improvement in prediction of the SDME, be it in the firing rate or the codeword distributions, can be directly ascribed to the effect of the interaction terms, β_{ij}.

We found that the SDME predicted the firing rate of single cells better than the LN models, with the normalized correlation coefficient between the true and predicted firing rate, $\mathrm{Corr}(r_i(t), r_i^{SDME}(t))$ being 0.74 ± 0.06 (mean ± std across 100 cells), as shown in Fig. 2b. The differences between the SDME and the LN models become more striking on the level of the activity patterns of the whole population. Figures 3a,b show the log-likelihood ratio for the population activity patterns $\mathbf{x} = \{x_i\}$ under the two models, showing that the SDME can be orders of magnitude better in predicting the population neural response. These differences are large in particular for those stimuli that elicit a strong response (Fig. 3c), that is, precisely where the response consists of synchronous spiking and the structure of the codewords can be nontrivial. Moreover, the difference between the models becomes increasingly significant with the size of the population N, and particularly apparent for groups of more than 20 cells (Fig. 3d).

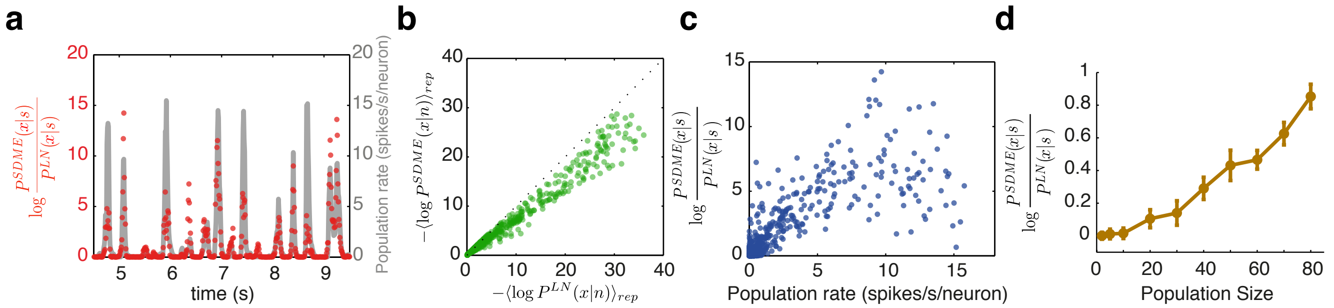

FIG. 3: **SDME model predicts population activity patterns for $N = 100$ neurons better than conditionally independent LN models.** **(a)** The log-likelihood ratio of the population firing patterns under the SDME model and under a collection of LN models, shown as a function of time (red) for an example stimulus repeat. For reference, the average population firing rate is shown in grey. **(b)** A scatterplot of the log-likelihoods under the SDME and LN models; each dot represents an average over activity patterns $\{x_i\}$ (across repeats) at a given time bin; dotted line shows equality. **(c)** The log-likelihood ratio under the SDME and LN models as a function of the average firing rate in the population; SDME outperforms LN models in particular for patterns with more spiking activity. **(d)** The average likelihood ratio between the SDME and LN models as a function of the population size, N (error bars = std over 10 randomly chosen groups of neurons at that N).

We next examined how well various models for the neural codebook, $P(\{x_i\}|\mathbf{s})$, explain the total vocabulary, that is, the distribution of neural codewords observed across the whole duration of the experiment, $P(\{x_i\}) = \langle P(\{x_i\}|\mathbf{s}(t))\rangle_t$. Despite the nominally large space of possible codewords—much larger than the total number of samples in the experiment ($2^N \gg T$)—the sparsity of spikes and the correlations between neurons restrict the vocabulary to a much smaller set of patterns. Some of these occur many times during our stimulus presentation, allowing us to estimate their empirical probability, $P^{data}(\{x_i\})$, directly from the experiment, and compare it to the model prediction [35]. The most prominent example of such frequently observed codewords is the silent pattern, $x_i = 0$, which is seen $\sim 72\%$ of the time.

Figure 4 shows the likelihood ratio of the model probability and empirical probability for various codewords observed in the experiment, as a function of the rate at which these codewords appear in the experiment. While SDME model in Fig. 4a does not predict the frequencies of all patterns perfectly, it strongly outperforms the uncoupled set of LN models in Fig. 4b, and has a slightly better performance than the full conditionally independent model (Fig. 4c), despite the fact that the latter is determined by $N \times 1000 = 1 \cdot 10^5$ parameters, the firing rates of every cell in every time bin. On average, SDME predicts the probabilities of the patterns of activity with no bias, and with a standard deviation of $\log(P^{SDME}/P^{data})$ of about 1; uncoupled LN models in comparison are biased and have a spread that is more than twice as large. Even more striking is the fact that LN models assign such low probabilities to some codewords that they are never generated during our Monte Carlo sampling (and are therefore not even shown in scatterplots of Fig. 4), while they are frequently observed in the experiment. This discrepancy is quantified by enumerating the M most probable patterns in the data and in the model (by sampling; see Methods), and measuring the size of the intersection of the two sets of patterns; in other words, we ask if the model is even able to access all the patterns that one is likely to record in the experiment. As shown in the third row of Fig. 4, SDME does well on this task, with 434 codewords in the intersection of the 500 most likely patterns in the data and the model; this is a much better performance than for the uncoupled model, and slightly better than for the conditionally independent model.

The SDME model was constructed to capture exactly the total correlations in neuronal spiking, $C_{ij} = \langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle$. With repeated stimulus, this total correlation can be broken down into the signal and noise components. The signal correlations, C^s_{ij}, are inferred by applying the same formula as for the total correlation, but on the spiking raster where the repeated trial indices have been randomly and independently permuted for each time bin. This removes any correlation due to interactions between spikes on simultaneously recorded trials, and only leaves the correlations induced by the response being locked to the stimulus. The noise correlation, C^n_{ij}, is then defined as the difference between the total and the signal components, $C^n_{ij} = C_{ij} - C^s_{ij}$. We calculated the noise correlations between all pairs in our $N = 100$ neuron dataset. By their definition, the conditionally independent models cannot reproduce C^n_{ij}, which are always zero. To assess the performance of the SDME, we drew samples from our model distribution using the Monte Carlo simulation and compared the noise correlations in the simulated rasters to the true noise correlations. The model prediction tightly correlates with the measured values, as shown in Fig. 5. We observe a systematic deviation of $\sim 25\%$, most likely because the assumed dependence on the stimulus through

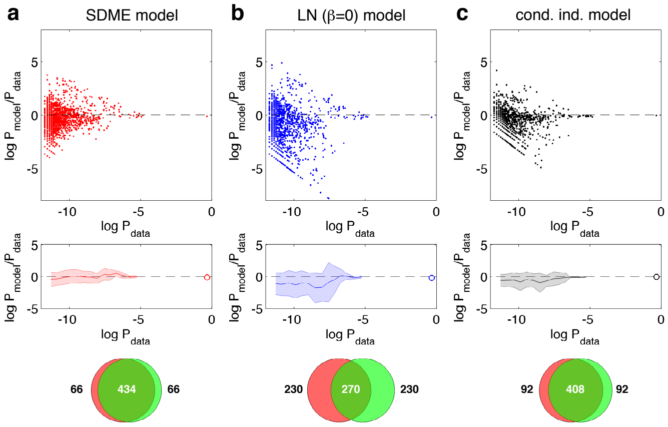

FIG. 4: **The performance of various models in accounting for the total vocabulary of the population, $P(\{x_i\})$.** The results for the SDME model are shown in **(a)**, the results for an uncoupled set of LN models in **(b)**, the results for a full conditionally independent model in **(c)**. The first row displays the log ratio of model to empirical probabilities for various codewords (dots), as a function of that codeword's empirical frequency in the recorded data. The model probabilities were estimated by generating Monte Carlo samples from the corresponding model distributions (see Methods); only patterns that were generated in the MC run as well as found in the recorded data are shown. The second row summarizes this scatterplot by binning codewords according to their frequency, and showing the average log probability ratio in the bin (solid line), as well as the 1 std scatter across the codewords in the bin (shaded area). The highly probable all-silent state, $\{x_i\} = 0$, is shown separately as a circle. The third row shows the overlap between 500 most frequent patterns in the data and 500 most likely patterns generated by the model (see text).

one linear filter per neuron is insufficient to capture the complete dependence on stimulus, thereby underestimating the full structure of stimulus correlation and inducing an excess in the noise correlation. Despite this, the degree of correspondence in noise correlations observed in Fig. 5 is telling us that SDME has clearly captured a large amount of noise covariance structure in neural firing.

How should we interpret the inferred parameters of the SDME model? LN models have a clear "mechanistic" interpretation in terms of the cell's receptive field and the nonlinear spiking mechanism. Here, similarly, the stimulus dependent part of the model for each cell, α_i, is a nonlinear function of a filtered version of the stimulus $g_i(t) = \mathbf{k}_i \cdot \mathbf{s}(t)$; in the absence of neuron-to-neuron couplings, the nonlinearity of every neuron would correspond to $\mathcal{N}_i(g_i) \sim f(\alpha_i(g_i))$, where $f(\cdot) = \exp(\cdot)/(1+\exp(\cdot))$, according to Eq. (4). The dependence of α_i on the stimulus projection g_i is similar across the recorded cells as shown in Fig. 6a; as expected, higher overlaps with the linear filter induce higher probability of spiking.

The pairwise interaction terms in the model, β_{ij}, are symmetric, static, and stimulus independent by construction.

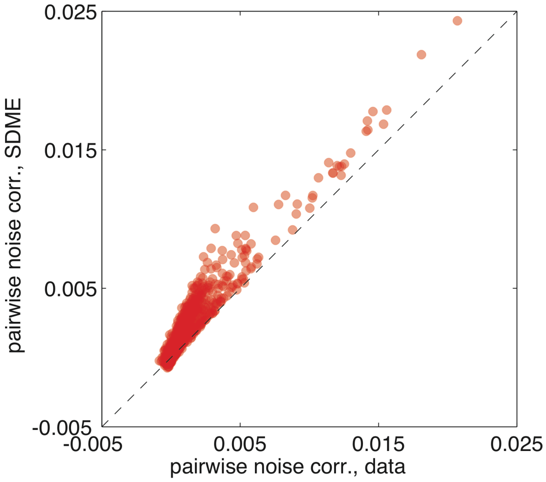

FIG. 5: **Measured vs predicted noise correlations for the SDME model.** Noise correlation (see text) is estimated from recorded data for every pair of neurons, and plotted against the noise correlation predicted by the SDME model (each pair of neurons = one dot; shown are $N(N-1)/2$ dots for $N = 100$ neurons). Conditionally independent models predict zero noise correlation for all pairs.

As such, they represent only functional and not physical (i.e. synaptic) connections between the cells. Fig. 6b shows the pairwise interaction map for 100 cells; the histogram of their values (in Fig. 6c) reflects that they can be of both signs, but the distribution has a stronger positive tail, i.e. a number of cell pairs tend to spike together or be silent together with a probability than is higher than expected from their respective LN models. We can compare these interactions to the interactions of a static (non-stimulus-dependent) maximum entropy model for the population vocabulary [26, 28]:

$$P^{ME}(\{x_i\}) = \frac{1}{Z_0} \exp\left(\sum_i \alpha_i^0 x_i + \frac{1}{2} \sum_{ij} \beta_{ij}^0 x_i x_j \right). \quad (5)$$

In this model for the total distribution of codewords, there is no stimulus dependence, and the parameters α_i^0 and β_{ij}^0 are chosen to that the distribution is as random as possible, while reproducing exactly the measured mean firing rate of every neuron $\langle x_i \rangle_{data} = \langle x_i \rangle_{PME}$, and every pairwise correlation, $\langle x_i x_j \rangle_{data} = \langle x_i x_j \rangle_{PME}$, across the whole duration of the experiment.

Interestingly, we find that the pairwise interaction terms in the SDME model of Eq. (3) are closely related to the interactions in the static pairwise maximum entropy model of Eq. (5): SDME interactions, β_{ij}, tend to be smaller in magnitude, but have an equal sign and relative ordering, as the static ME interactions, β_{ij}^0. Some degree of correspondence is expected: an interaction between neurons i and j in the static ME model captures the combined effect of the stimulus and noise correlations,

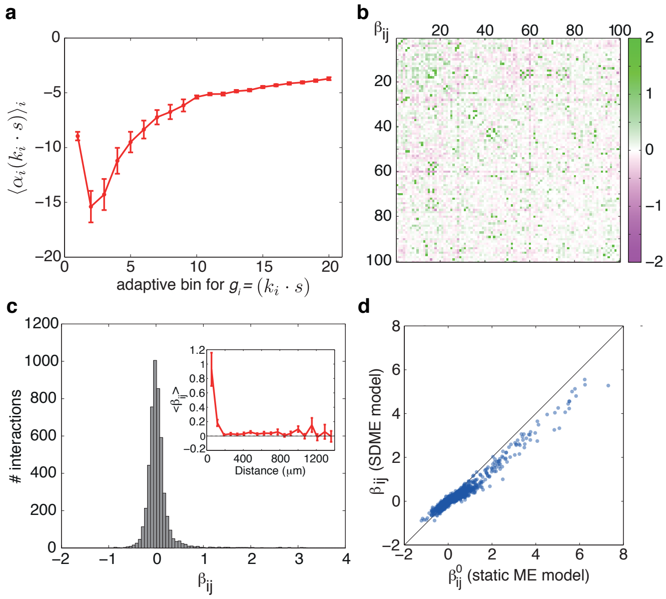

FIG. 6: **SDME model parameters.** **(a)** Average values of the LN-like driving term, $\alpha_i(g_i)$, where $g_i = \mathbf{k}_i \cdot \mathbf{s}$, across all cells i (error bars = std across cells), for each of the $K = 20$ adaptive bins for g_i (see Methods). **(b)** Pairwise interaction map β_{ij} of the SDME model, between all $N = 100$ neurons in the experiment. **(c)** Histogram of pairwise interaction values from (b), and their average value as a function of the distance between cells (inset). **(d)** For each pair of cells i and j, we plot the value of β^0_{ij} under the static maximum entropy model of Eq. (5) vs. the β_{ij} from the stimulus-dependent maximum entropy model of Eq. (3).

while in the corresponding SDME interaction, (most of) the stimulus correlation has been factored out into the correlated dynamics of the inputs to the neurons i and j, i.e. $\alpha_i(g_i(t))$ and $\alpha_j(g_j(t))$. The surprisingly high degree of correspondence, however, indicates that even the interactions learned from static maximum entropy models can account for, up to a scaling factor, the pairwise neuron dependencies that are *not* due to the correlated stimulus inputs.

The SDME model is an approximation to the neural codebook, $P(\{x_i\}|\mathbf{s})$, while the static ME model describes the population vocabulary, $P(\{x_i\})$. With these two distributions in hand, we can explore how the population jointly encodes the information about the stimulus into neural codewords—the joint activity patterns of spiking and silence. We make use of the fact that we can estimate the entropy of the maximum entropy distributions using a procedure of heat capacity integration, as explained in Refs. [27, 32] (see Methods). Information (in bits) per codeword is

$$\begin{aligned} I(\{x_i\};\mathbf{s}) &= \int d\mathbf{s}\, P(\mathbf{s}) \sum_{\{x_i\}} P(\{x_i\}|\mathbf{s}) \log_2 \frac{P(\{x_i\}|\mathbf{s})}{P(\{x_i\})} \\ &= S[P(\{x_i\})] - \langle S[P(\{x_i\}|\mathbf{s})]\rangle_{P(\mathbf{s})}; \qquad (6) \end{aligned}$$

that is, the information can be written as a difference of the entropy of the neural vocabulary, and the noise entropy (the average of the entropy of the codebook), where the entropy is $S[p(x)] = -\int dx\, p(x) \log_2 p(x)$. Because of the maximum entropy property of our model for $P^{ME}(\{x_i\})$, the entropy of our static pairwise model in Eq. (5) is an upper bound on the transmitted information; expressed as an entropy rate, this amounts to $s \equiv S[P^{ME}(\{x_i\})]/\Delta t \approx 730$ bit/s.

The brain does not have direct access to the stimulus, but only receives codewords $\{x_i\}$, "drawn" from $P(\{x_i\})$, by the retina. At every moment in time, $-\log_2 P(\{x_i\})$ measures the *surprise* about the output of the retina, and thus about the stimulus. We, as experimenters—but not the brain—have access to stimulus repeats and thus to $P(\{x_i\}|\mathbf{s}(t))$, so we can compute the average value of surprise (per unit time) at every instant t in the stimulus:

$$\mathcal{S}(t) = -\frac{1}{\Delta t} \sum_{\{x_i\}} P(\{x_i\}|\mathbf{s}(t)) \log_2 P(\{x_i\}). \qquad (7)$$

This quantity can be expressed using the entropies and the learned parameters of our maximum entropy models, and is plotted as a function of time in Fig. 7. Since averaging across time is equal to averaging over the stimulus ensemble, we see from Eq. (7) that $\langle\mathcal{S}(t)\rangle_t$ would have to be identically equal to $S[P(\{x\})]$ under the condition that $\langle P(\{x_i\}|\mathbf{s}(t))\rangle_t = P(\{x_i\})$ (marginalization). Since we build models for $P(\{x_i\})$ (static ME) and $P(\{x_i\}|\mathbf{s})$ (SDME) from data independently, they need not obey the marginalization condition exactly, but they will do so if they provide a good account of the data. Indeed, by using the static ME and SDME distributions in Eq. (7) for surprise, we find that $\langle\mathcal{S}(t)\rangle_t \approx 740$ bit/s, very close to the entropy rate s of the total vocabulary and within our estimated 1% error bars on entropy computation.

To estimate the information transmission, we have to subtract the noise entropy rate from the output entropy rate s, as dictated by Eq. (6). The entropy of the SDME model is an upper bound on the noise entropy; since this is not a lower bound, we cannot put a strict bound on the information transmission, but can nevertheless estimate it. Figure 7 shows the "instantaneous information", $\mathcal{I}(t) = \mathcal{S}(t) - S[P^{SDME}(\{x_i\}|\mathbf{s}(t))]/\Delta t$, as a function of time; from Eq. (6), the mutual information rate is a time average of this quantity, $R = I(\{x_i\};\mathbf{s})/\Delta t = \langle\mathcal{I}(t)\rangle_t$. We find $R \approx 130$ bit/s. This quantity can be compared to the total entropy rate of the stimulus itself (which must be higher than R), which in our case is ≈ 210 bit/s (see Methods). While our estimates seem to indicate that a lot of vocabulary bandwidth (730 bit/s) is "lost" to noise (600 bit/s), the last comparison shows that the Gaussian FFF stimulus source itself is not very rich, so that the estimated information transmission takes up more than half of the actual entropy rate of the source.

Lastly, we asked how important is the inclusion of

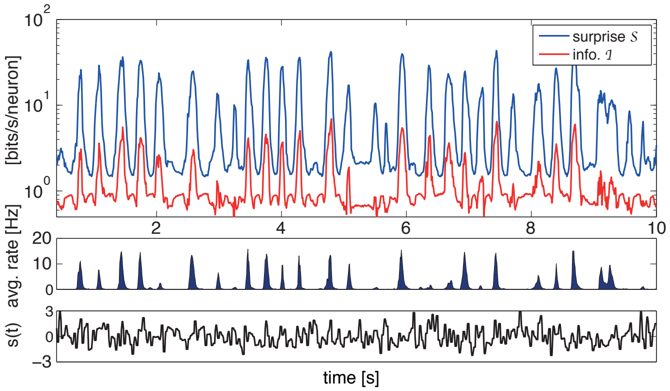

FIG. 7: **Surprise and information transmission estimated from the SDME model. (a)** Surprise rate (blue) is estimated from the static ME and SDME models assuming independence of codewords across time bins. The instantaneous information rate (red) is the difference between the surprise and the noise entropy rate, estimated from the SDME model (see text). The information transmission rate is the average of the instantaneous information across time. **(b)** Population firing rate as a function of time shows that bursts of spiking strongly correlate with the bursts of surprise and information transmission in the population. **(c)** The stimulus (normalized to zero mean and unit variance) is shown for reference as a function of time.

pairwise interactions, β_{ij}, into the SDME model, compared to a set of uncoupled LN models, when accounting for information transmission. We therefore estimated the noise entropy rate for a set of uncoupled LN models, $S[P^{LN}(\{x_i\}|\mathbf{s}(t))]/\Delta t$, which was found to be ≈ 770 bit/s, considerably higher than the noise entropy of the SDME model. Crucially, this noise entropy rate is larger than the total entropy rate s estimated above, which is impossible for consistent models of the neural codebook and the vocabulary (since it would lead to negative information rates). This failure is a quantitative demonstration of the inability of the uncoupled LN models to reproduce the statistics of the population vocabulary, as shown in Fig. 4b, despite a seemingly small performance difference on the level of single cell PSTH prediction.

DISCUSSION

We presented a modeling framework for stimulus encoding by large populations of neurons, which combines an individual neuronal receptive field model, with the ability to include pairwise interactions between neurons. The result is a stimulus-dependent maximum entropy (SDME) model, which is the most parsimonious model of the population response to the stimulus that reproduces the linear-nonlinear (LN) aspect of single cells, as well as the correlation structure between neurons. In two limiting cases, the SDME model reduces to known models: if the single cell parameters α are static, SDME becomes the static maximum entropy model of the population vocabulary; if the couplings β are 0, SDME becomes a set of uncoupled LN models. The framework is general, and could be easily applied to other neural systems.

We applied this modeling framework to the salamander retina presented with Gaussian white noise stimuli, and found that the interactions between neurons play an important role in determining the detailed patterns of population response. In particular, the SDME model gave better prediction of PSTH of single cells, yielded orders of magnitude improvement in describing the population patterns, and captured significant aspects of noise correlations. The deviations between the SDME and the uncoupled LN model became significant for > 20 cells, and tended to occur at "interesting" times in the stimulus, precisely when the neural population was not silent.

The responses of the neural system in the maximum entropy framework are binary codewords of spiking and silence across the neural population. The choice of timescale over which these codewords are defined, here $\Delta t = 10$ ms, is short enough such that multiple spikes are rarely observed in the same time bin, but long enough so that most of the strong spike-spike interactions (as well as fine temporal detail, such as spike-timing jitter) occur within a single bin. This allows us to view successive time bins as codewords, although some statistical dependence between them remains, possibly in the conditional SDME model (due to multiple timescales on which the neurons respond to stimuli), and certainly in the static ME model [31]. If we were to make the time scale much shorter, e.g. by an order of magnitude, we could make the conditional independence assumption of the responses given the stimuli *and* previous spiking, which would lead us to GLM models [40] or nonequilibrium generalizations of Ising models [47]. GLMs, in particular, are excellent generating models for precise spiking rasters, are easy to infer, and allow for asymmetric couplings between neurons. However, the inference in all these cases is tractable because there are no interactions between the spikes *within* the same time bin (as there are in SDME). This necessitates the use of very short time bins and introduces strong dependencies between successive time bins, making the interpretation of the discretized neural responses in terms of individual codewords difficult. For this reason, GLM and SDME are complementary approaches: the first allows for a temporally-detailed probabilistic description of a spiking process, while the second gives an explicit expression for the probability distribution over codewords in longer temporal bins.

SDME allowed us to improve over LN models for salamander retinal ganglion cells both in terms of the PSTH prediction and, especially, in terms of population activity patterns. Interestingly, for parasol cells in the macaque retina under flickering checkerboard stimulation, the generalized linear model did not yield firing rate improvement relative to uncoupled LN models (but did improve higher order statistics of firing) [40]. In both cases, however, the improvements reflect the role of dependencies among cells in encoding the stimulus, and their effect becomes apparent when we ask questions about information transmission by a neural population. Maximum entropy models can only put an upper bounds on the total entropy and the noise entropy of the neural code (and this statement remains true even if successive codewords are not independent), and as such cannot set a strict bound, but only give an estimate, for the information transmission. Nevertheless, ignoring the inter-neuron dependencies and using an uncoupled set of LN models predicts the total population responses so badly that the estimated noise entropy is higher than the upper bound on the total entropy, which is a clear impossibility, while the SDME model gives transmission rates that appear reasonable (and positive), amounting to about 60% of the source entropy rate.

Tkačik and colleagues [61] have suggested that one can interpret β_{ij} in an SDME model as a prior over the activity patterns that the population would use to optimally encode the stimulus. For low noise level they argued that the prior should be minimal (and could help decorrelate the responses), as the population could faithfully encode the stimulus, whereas in the noisy regime, the prior should match the statistics of the sensory world and thus counteract the effects of noise. Similarly, Berkes and colleagues [62] suggested a similar reason for the similarity of ongoing and induced activity patterns in the visual cortex. Our results show that interactions are necessary for capturing the network encoding, and implicitly reflect the existence of such a prior. The recovered interactions are strongly correlated with the interaction parameters of a static, stimulus independent model over the distribution of patterns, making it possible for the brain (which only has access to the spikes, not the stimulus) to learn these values. Whether the interactions are matched to the statistics of the visual inputs as suggested by Ref [61] is the focus of future work. In parallel, increasingly detailed statistical models of neural codes going beyond SDME (e.g. by including temporal dependencies as in Ref [48]), and efforts to infer such models from experimental data, should focus our attention on population-level statistics and on finding principled information-theoretic measures for quantifying the code, like the surprise and instantaneous information suggested here.

METHODS

Electrophysiology. Experiments were performed on the adult tiger salamander, *Ambystoma tigrinum.* All experiments were in accordance with Ben-Gurion University of the Negev and government regulations. Extracted retinas were placed with the ganglion cell layer facing a multielectrode array with 252 electrodes (Ayanda Biosystems, Switzerland), and superfused with oxygenated Ringer medium at room temperature. Extracellularly recorded signals were amplified (MultiChannel Systems, Germany) and digitized at 10k Samples/s, and spike-sorted using custom software written in MATLAB.

Visual stimulation. Stimuli were projected onto the retina from a CRT video monitor (ViewSonic G90fB) at a frame rate of 60 Hz; each movie frame was presented twice, using standard optics. Full Field Flicker (FFF) stimuli were generated by independently sampling spatially uniform gray levels (with a resolution of 8 bits) from a Gaussian distribution, with mean luminance of 147 lux and the standard deviation of 33 lux. These data allow us to estimate the entropy rate of the source (as used in the main text), by multiplying the entropy of the luminance distribution with the refresh rate. To estimate the cells' receptive fields, checkerboard stimulus was generated by selecting each checker ($\sim 100\ \mu$m on the retina) randomly every 33 ms to be either black or white. To identify the RF centers, a two-dimensional Gaussian was fitted to the spatial profile of the response. The movies were gamma corrected for the computer monitor. In all cases the visual stimulus entirely covered the retinal patch that was used for the experiment.

Inferring SDME from data. The LN model for each neuron i consists of the linear filter $\mathbf{k}_i$, and the nonlinear function $\mathcal{N}_i$, which is defined pointwise on a set of binned values for the generator signal, $g_i = \mathbf{k}_i \cdot \mathbf{s}$. We used binning into $K = 20$ bins such that initially each bin contains roughly the same number of values for g_i, but subsequently the binning is adaptively adjusted (separately for each neuron) to be denser at higher values of g_i, where the firing rates are higher. We fitted LN models with varying number of K bins, and have chosen $K = 20$ when the performance of the LN models appeared to saturate [63].

To find the parameters of the stimulus-dependent maximum entropy model $(\alpha_i(g_i), \beta_{ij})$, we retained the binning of the generator signal used for LN model construction. Given trial values for the SDME parameters, we estimated the chosen expectation values (covariance matrix C_{ij} in firing, and the firing rate conditional on g_i, $r_i(g_i)$) by Monte Carlo sampling from the trial distribution in Eq. (3); the learning step of the algorithm is computed by comparing the expectation values in the trial distribution and the empirical distribution (computed over the training half of the stimulus repeats). In detail, we used a gradient ascent algorithm, applying a combination of Gibbs sampling and importance sampling in order to efficiently estimate the gradient, by using optimizations similar to those described in Ref. [64]. Sampling was carried out in parallel on a 16 node cluster with two 2.66GHz Intel Quad-Core Xeon processors and 16GB of memory per node. The calculation was terminated when the average error in firing rates and coincident firing rates reached below 1% and 5% respectively, which is within the experimental error.

To compute the single neuron PSTH and compare the distributions of codewords from the model to the empirical distribution, we used Metropolis Monte Carlo sampling to draw codewords from the model distributions; we drew 5000 independent samples (to draw uncorrelated configurations, a sample was recorded only after 100 "spin-flip" trials) for every timepoint, for a total of $5 \cdot 10^6$ samples; the same procedure was used also to draw from the uncoupled ($\beta = 0$) models. To estimate the entropies of high dimensional SDME distributions, we used the "heat capacity integration" method, detailed in Ref [32]. Briefly, a maximum entropy model $P(\mathbf{x}) = Z^{-1} \exp(-E(\mathbf{x}))$ (where E is the Hamiltonian function determined by the choice of constrained operators and the conjugated parameters) is extended by introducing a new parameter T, much like the temperature in physics, so that $P_T(\mathbf{x}) = Z_T^{-1} \exp(-E(\mathbf{x})/T)$. The entropy of the distribution is given by $S[P_{T=1}] = \int_0^1 C(T)/T dT$, where the heat capacity $C(T) = \sigma_E^2(T)/T^2$, and the variance in energy can be estimated at each T by Monte Carlo sampling. In practice, we run a separate Monte Carlo sampling for a finely discretized interval of temperatures, $T \in [0, 1]$, estimate $C(T)$ for each temperature, and numerically integrate to get the entropy S. We have previously shown that this procedure yields robust entropy estimates even for large numbers of neurons [27, 32].

Acknowledgements. This work was supported by The Israel Science Foundation and the Human Frontiers Science Program.

[1] Rieke F, Warland D, de Ruyter van Steveninck RR & Bialek W (1997) *Spikes: Exploring the Neural Code.* MIT Press, Cambridge, MA.

[2] Agüera y Arcas B & Fairhall AL (2003) What causes a neuron to spike? *Neural Comput* **15:** 1789–1807.

[3] Bialek W & de Ruyter van Steveninck RR (2005) Features and dimensions: Motion estimation in fly vision. *arXiv.org:*q-bio/0505003.

[4] Schwartz O, Pillow JW, Rust NC & Simoncelli EP (2006) Spike-triggered neural characterization. *J Vis* **6:** 484–507.

[5] Stopfer M, Bhagavan S, Smith BH, Laurent G (1997) Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* 390:70–4.

[6] Riehle A, Grün S, Diesmann M, Aertsen A (1997) Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278:1950–3.

[7] Harris KD, Csicsvari J, Hirase H, Dragoi G, Buzsáki G (2003) Organization of cell assemblies in the hippocampus. *Nature* 424:552–6.

[8] Averbeck BB, Lee D (2004) Coding and transmission of information by neural ensembles. *Trends Neurosci* 27:225–30.

[9] Brunel N, Nadal JP (1998) Mutual information, Fisher information, and population coding. *Neural Comp* 10:1731–1757.

[10] Abbott LF, Dayan P (1998) he Effect of Correlated Variability on the Accuracy of a Population Code. *Neural Comp.* 11:91-102

[11] Sompolinsky H, Yoon H, KAng K, Shamir M (2001) Population coding in neuronal systems with correlated noise. *Phys Rev E* 64:8095–8100.

[12] Schneidman E, Bialek W, Berry MJ (2003) Synergy, redundancy, and independence in population codes. *J Neurosci* 23:11539–53.

[13] Pola, G, Thiele A, Hoffmann K-P, Panzeri, S (2003) An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network: Comput. Neural Syst.* 14:35–60.

[14] Nirenberg S, Latham PE (2003) Decoding neuronal spike trains: How important are correlations?. *Proc. Natl. Acad. Sci. USA* 100:7348–7353.

[15] Averbeck B, Latham PR, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7:358–366 .

[16] Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area mt: time scales and relationship to behavior. *Journal of Neuroscience* 21:1676–97.

[17] Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* **327:** 584–7.

[18] Puchalla JL, Schneidman E, Harris RA & Berry MJ 2nd (2005) Redundancy in the population code of the retina. *Neuron* **46:** 493–504.

[19] Narayanan NS, Kimchi EY, Laubach M (2005) Redundancy and synergy of neuronal ensembles in motor cortex. *J Neurosci* 25:4207–16.

[20] Chechik G, et al. (2006) Reduction of information redundancy in the ascending auditory pathway. *Neuron* 51:359–68.

[21] Nirenberg S, Carcieri SM, Jacobs AL & Latham PE (2001) Retinal ganglion cells act largely as independent encoders. *Nature* **411:** 698–701.

[22] HB Barlow, Possible principles underlying the transformation of sensory messages. *Sensory communication,* ed Rosenblith W (MIT Press, Cambridge, MA), pp 217–234 (1961).

[23] JJ Atick & AN Redlich, Towards a theory of early visual processing. *Neural Comp* **2,** 308–320 (1990).

[24] Barlow H (2001) Redundancy reduction revisited. *Network* 12:241–53.

[25] Schnitzer MJ, Meister M (2003) Multineuronal firing patterns in the signal from eye to brain. *Neuron* 37:499–511.

[26] Schneidman E, Berry MJ 2nd, Segev R & Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440:** 1007–12.

[27] G Tkačik, E Schneidman, MJ Berry II & W Bialek (2006) Ising models for networks of real neurons. arXiv:q-bio/0611072.

[28] J Shlens, GD Field, JL Gaulthier, MI Grivich, D Petrusca, A Sher, AM Litke & EJ Chichilnisky (2006) The structure of multi-neuron firing patterns in primate retina. *J Neurosci* **26:** 8254-66.

[29] A Tang, D Jackson, J Hobbs, W Chen, JL Smith, H PAtel, A Prieto, D Petruscam MI Grivich, A Sher, P Hottowy, W Dabrowski, AM Litke & JM Beggs (2008) A

maximum entropy model applied to spatial and temporal correlations from cortical networks *in vitro. J Neurosci* **28:** 505–518.

[30] J Shlens, GD Field, JL Gaulthier, M Greschner, A Sher, AM Litke & EJ Chichilnisky (2009) The structure of large-scale synchronized firing in primate retina. *J Neurosci* **29:** 5022–31.

[31] O Marre, SE Boustani, Y Fregnac & A Destexhe (2009) Prediction of spatio–temporal patterns of neural activity from pairwise correlations. *Phys Rev Lett* **102:** 138101.

[32] G Tkačik, E Schneidman, MJ Berry II & W Bialek (2009) Spin-glass models for a network of real neurons. arXiv:0912.5409 (2009).

[33] Ganmor E, Segev R, Schneidman E (2011) The architecture of functional interaction networks in the retina. *J Neurosci* 31:3044–54.

[34] Ohiorhenuan IE, Mechler F, Purpura KP, Schmid AM, Hu Q & Victor JD (2010) Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* **466:** 617–21.

[35] Ganmor E, Segev R & Schneidman E (2011) Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proc Nat'l Acad Sci USA* **108:** 9679–84.

[36] Warland DK, Reinagel P & Meister M (1997) Decoding visual information from a population of retinal ganglion cells. *J Neurophys* **78:** 2336–2350.

[37] Dan Y, Alonso JM, Usrey WM, Reid RC (1998) Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nat Neurosci* 1:501–7.

[38] Hatsopoulos NG, Ojakangas CL, Paninski L, Donoghue JP (1998) Information about movement direction obtained from synchronous activity of motor cortical neurons. *Proc Natl Acad Sci USA* 95:15706–11.

[39] Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J Neurosci* 18:7411–25.

[40] Pillow JW, Shlens J, Paninski L, Shear A, Litke AM, Chichilnisky EJ & Simoncelli EP (2008) Spatio-temporal correlations and visual signaling in a complete neural population. *Nature* **454:** 995–9.

[41] E Schneidman, S Still, MJ Berry II & W Bialek (2003) Network information and connected correlations. *Phys Rev Lett* **91:** 238701.

[42] S Cocco, S Leibler & R Monasson (2009) Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proc Nat'l Acad Sci USA* **106:** 14058–62.

[43] S Cocco & R Monasson (2011) Adaptive cluster expansion for inferring Boltzmann machines with noisy data. *Phys Rev Lett* **106:** 090601.

[44] Y Roudi, E Aurell & JA Hertz (2009) Statistical physics of pairwise probability models. *Front Comput Neurosci* **3:** 22.

[45] Y Roudi, S Nirenberg & PE Latham (2009) Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS Comput Biol* **5:** e1000380.

[46] Y Roudi, J Trycha & J Hertz (2009) The ising model for neural data: model quality and approximate methods for extracting functional connectivity. *Phys Rev E* **79:** 051915.

[47] Y Roudi & J Hertz (2011) Mean field theory for nonequilibrium network reconstruction. *Phys Rev Lett* **106:** 048702.

[48] Vasquez JC, Marre O, Palacios AG, Berry MJ 2nd & Cessac B (2011) Gibbs distribution analysis of temporal correlation structure in retina ganglion cells. *J Physiol Paris*, *in press.*

[49] Macke JH, Opper M & Bethge M (2011) Common input explains higher-order correlations and entropy in a simple model of neural population activity. *Phys Rev Lett* **106:** 208102.

[50] M Mezard & T Mora, Constraint satisfaction problems and neural networks: a statistical physics perspective (2009) *J Physiol Paris* **103:** 107–113.

[51] B Cessac, H Rostro, JC Vasques & T Vieville (2009) How Gibbs distributions may naturally arise from synaptic adaptation mechanisms. *J Stat Phys* **136:** 565–602.

[52] V Sessak & R Monasson (2009) Small-correlation expansions for the inverse Ising problem. *J Phys A* **42,** 055001.

[53] Segev R, Goodhouse J, Puchalla J & Berry MJ 2nd (2004) Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nat Neurosci* **7:** 1154–61.

[54] AL Fairhall, CA Burlingame, R Narasimhan, RA Harris, JL Puchalla & MJ Berry II (2006) Selectivity for multiple stimulus features in retinal ganglion cells. *J Neurophysiol* **96:** 2724–2738.

[55] G Tkačik, A Ghosh, E Schneidman & R Segev (2012) Retinal adaptation and invariance to changes in higher-order stimulus statistics. *arXiv.org:1201.3552.*

[56] J Keat, P Reinagel, RC Reid & M Meister (2001) Predicting every spike: a model for the responses of visual neurons. *Neuron* **30:** 803–817.

[57] Ozuysal Y & Baccus SA (2012) Linking the computational structure of variance adaptation to biophysical mechanisms. *Neuron* **73:** 1002-1015.

[58] Tkačik G, Information flow in biological networks. *Thesis*, Princeton University, 2007.

[59] Granot-Atdegi E, Tkačik G, Segev R & Schneidman E (2010) A stimulus-dependent maximum entropy model of the retinal population neural code. *Front Neurosci Conference Abstract* COSYNE 2010.

[60] ET Jaynes (1957) Information theory and statistical mechanics. *Phys Rev* **106:** 620–630.

[61] Tkačik G, Prentice JS, Balasubramanian V & Schneidman E (2010) Optimal population coding by noisy spiking neurons. *Proc Nat'l Acad Sci USA* **107:** 14419–14424.

[62] Berkes P, Orban G, Lengyel M & Fiser J (2011) Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* **331:** 83–7.

[63] E Granot-Atedgi (2009) Stimulus-dependent maximum entropy models and decoding of naturalistic movies from large populations of retinal neurons. *Thesis.* Weizmann Institute of Science, Israel.

[64] T Broderick, M Dudik, G Tkačik, RE Schapire & W Bialek (2007) Faster solutions of the inverse pairwise Ising problem. arXiv:0712.2437.